(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2001 (27.09.2001)

PCT

(10) International Publication Number
WO 01/71042 A2

c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). LI, Peter, W., D.; c/o Celera, 45 West Gude Drive. Rockville, MD 20850 (US). MYERS, Eugene, W.: c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US).
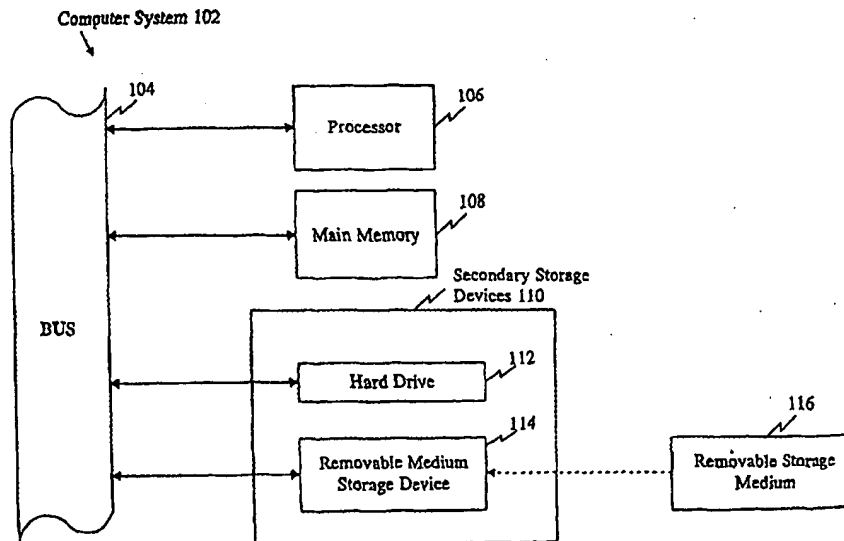
(74) Agent: CELERA GENOMICS CORP.; Robert A. Millman, 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

[Continued on next page]

(54) Title: DETECTION KITS, SUCH AS NUCLEIC ACID ARRAYS, FOR DETECTING THE EXPRESSION OF 10,000 OR MORE DROSOPHILA GENES AND USES THEREOF

(57) Abstract: The present invention is based on the sequencing and assembly of the *Drosophila melanogaster* genome. The present invention provides the primary nucleotide sequence of a large portion of the *Drosophila melanogaster* genome in a series of genomic and predicted transcript sequences. This information is provided in the form of genomic, transcript and protein sequence information and can be used to generate nucleic acid detection reagents and kits such a nucleic acid arrays.

WO 01/71042 A2

WO 01/71042 A2

Untitled

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 17, 2004, 07:14:52 ; Search time 131 Seconds
                                  (without alignments)
                                  5786.221 Million cell updates/sec

Title: PCT-US03-24982A-15
Perfect score: 11007
Sequence: 1 MTASDKYTYQRTVLCLARVL.........QNFRTRTYDMIQYYQNQIPY 2113

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_23Sep04:*
                 1: geneseqp1980s:*
                 2: geneseqp1990s:*
                 3: geneseqp2000s:*
                 4: geneseqp2001s:*
                 5: geneseqp2002s:*
                 6: geneseqp2003as:*
                 7: geneseqp2003bs:*
                 8: geneseqp2004s:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                          SUMMARIES
                  %
                  Query
Result   Match
  No.   Score Match Length DB ID   Description
----
    1   11007 100.0   2113 4   ABB64885   Abb64885 Drosophil

                          ALIGNMENTS

RESULT 1
ABB64885
ID   ABB64885 standard; protein; 2113 AA.
XX
AC   ABB64885;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 21447.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.

```
XX
PD    27-SEP-2001.
XX
PF    23-MAR-2001; 2001WO-US009231.
XX
PR    23-MAR-2000; 2000US-0191637P.
PR    11-JUL-2000; 2000US-00614150.
XX
PA    (PEKE ) PE CORP NY.
XX
PI    Venter JC, Adams M, Li PWD, Myers EW;
XX
DR    WPI; 2001-656860/75.
DR    N-PSDB; ABL08988.
XX
PT    New isolated nucleic acid detection reagent for detecting 1000 or more
PT    genes from Drosophila and for elucidating cell signaling and cell-cell
PT    interactions.
XX
PS    Disclosure; SEQ ID NO 21447; 21pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated nucleic acid detection reagent
CC    capable of detecting 1000 or more genes from Drosophila. The invention is
CC    useful in developmental biology and in elucidating cell signalling and
CC    cell-cell interactions in higher eukaryotes for the development of
CC    insecticides, therapeutics and pharmaceutical drugs. The invention
CC    discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC    sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC    ABB72072). The sequence data for this patent did not form part of the
CC    printed specification, but was obtained in electronic format directly
CC    from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 2113 AA;
```

```
Query Match              100.0%;  Score 11007;  DB 4;  Length 2113;
Best Local Similarity    100.0%;  Pred. No. 0;
Matches 2113;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 MTASDKYTYQRTVLCLARVLAGIQPTPWDKVHVHLATHSTSHSASPPGNRLFALTLSGGH 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTASDKYTYQRTVLCLARVLAGIQPTPWDKVHVHLATHSTSHSASPPGNRLFALTLSGGH 60

Qy     61 KTDMRGRSGGGGKAVQTLFRYCPQENAAGVFCLDTRAQDAVIALGIYFLEGGCQHEGQIV 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 KTDMRGRSGGGGKAVQTLFRYCPQENAAGVFCLDTRAQDAVIALGIYFLEGGCQHEGQIV 120

Qy    121 PYLLRLAKCLPKAVWIDDARSNKVERVRIPSAEKFSFCLNTLLSDIAAKCPDSREEIILN 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 PYLLRLAKCLPKAVWIDDARSNKVERVRIPSAEKFSFCLNTLLSDIAAKCPDSREEIILN 180

Qy    181 QVETLSALANIVKSSRDSSSAPPPIILCKATVPLLFGLARSMGRYASNDPPLLCRIFPPE 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 QVETLSALANIVKSSRDSSSAPPPIILCKATVPLLFGLARSMGRYASNDPPLLCRIFPPE 240

Qy    241 LLPIQKGGGRDGTGSSSSASGTCGGSFSSSERLAATHHFRPIIPRSMSGSLAQAQNASYD 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 LLPIQKGGGRDGTGSSSSASGTCGGSFSSSERLAATHHFRPIIPRSMSGSLAQAQNASYD 300

Qy    301 DGRQRCAGGKPSKPSLHSYFSVPYDPRTHFFTRYGSSFNQFPNMRVCESPTKGGPRPLYR 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 DGRQRCAGGKPSKPSLHSYFSVPYDPRTHFFTRYGSSFNQFPNMRVCESPTKGGPRPLYR 360

Qy    361 VPPFPIQHLQTIFAVSKKLLTKDTLEHLDEQASDIFSLHQIKGYCYKSFSETLNLVLVTL 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 VPPFPIQHLQTIFAVSKKLLTKDTLEHLDEQASDIFSLHQIKGYCYKSFSETLNLVLVTL 420

Qy    421 LRELLQHQVDLPTPFTKDVQEFVKRLFLNGQTELQNKQQDQERERREENGIAVVNKYKVN 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 LRELLQHQVDLPTPFTKDVQEFVKRLFLNGQTELQNKQQDQERERREENGIAVVNKYKVN 480

Qy    481 VMANAACVDLLVWAIRDETEADKLCGRLSQKLNLELSHKIVMDHMPLLMVCLEGLGKLAH 540
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  VMANAACVDLLVWAIRDETEADKLCGRLSQKLNLELSHKIVMDHMPLLMVCLEGLGKLAH 540

Qy      541  KFPNIAGTSISYLRDFLVAPSPILGKLHDHAMQSLAQQKKEKELTPFKIAVQHSDSRTAV 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541  KFPNIAGTSISYLRDFLVAPSPILGKLHDHAMQSLAQQKKEKELTPFKIAVQHSDSRTAV 600

Qy      601  VIYGDNQKPPGSGTGRSGHAAFESLRDAAIENLSIALRAAHTLDQFCVPALVANVSNRLL 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601  VIYGDNQKPPGSGTGRSGHAAFESLRDAAIENLSIALRAAHTLDQFCVPALVANVSNRLL 660

Qy      661  QPKVWQMIISQCETHVFDEIMKMFSRVTVQSASLAYTSDPEHRKQFHHVSDAVVNALGNI 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661  QPKVWQMIISQCETHVFDEIMKMFSRVTVQSASLAYTSDPEHRKQFHHVSDAVVNALGNI 720

Qy      721  AANIQGDAEMLELLGKLLELFVQIGLDGERSYDNTPGAQKASSRAGNLGMLIPVIAVLVR 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721  AANIQGDAEMLELLGKLLELFVQIGLDGERSYDNTPGAQKASSRAGNLGMLIPVIAVLVR 780

Qy      781  RLPPIKNPRQRLHKLFKDFWAYCVVMGFTNARLWPADWYQGVQQIAAKSPLLISQTAHKS 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781  RLPPIKNPRQRLHKLFKDFWAYCVVMGFTNARLWPADWYQGVQQIAAKSPLLISQTAHKS 840

Qy      841  DMRELNYTLAIKSDSVNELRSQILVLLEHSSDNVATAINKLSFAQCTYLLSVYWLEMLRV 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841  DMRELNYTLAIKSDSVNELRSQILVLLEHSSDNVATAINKLSFAQCTYLLSVYWLEMLRV 900

Qy      901  ENADEPSLEPIMSYLCDTALQRDKTGIWQCVKCVADQVFEKFRNVLYAHDEIREKVLESQ 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901  ENADEPSLEPIMSYLCDTALQRDKTGIWQCVKCVADQVFEKFRNVLYAHDEIREKVLESQ 960

Qy      961  ATLLLVYFNHIHKPIQMVADQYLSFLVDRFPHLLWNRRVLWCMLDILQLLAYSLSLDPNE 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961  ATLLLVYFNHIHKPIQMVADQYLSFLVDRFPHLLWNRRVLWCMLDILQLLAYSLSLDPNE 1020

Qy     1021  ETPTLRVVSTPYTLQLMDSLPARELRLKDFADRCQGIVNEAMKWAPRSTRSHLQEYPNQI 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021  ETPTLRVVSTPYTLQLMDSLPARELRLKDFADRCQGIVNEAMKWAPRSTRSHLQEYPNQI 1080

Qy     1081  PTPVLAHHSGLALAFDSVVSSSAQHTGTMSKRPSCVNSDTPRFVSVLCLRSKYAGEISGL 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081  PTPVLAHHSGLALAFDSVVSSSAQHTGTMSKRPSCVNSDTPRFVSVLCLRSKYAGEISGL 1140

Qy     1141  LSVLSEKDKAGLADRLVSDVWEACAEKSDARHRGALWRATAYLIICSEISRKLLHAVASS 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141  LSVLSEKDKAGLADRLVSDVWEACAEKSDARHRGALWRATAYLIICSEISRKLLHAVASS 1200

Qy     1201  QLELFTESAMETAVECWQWVLTARQDLELCFIQEMVSAWQTTFEKRMGLFAWETEVTHPL 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201  QLELFTESAMETAVECWQWVLTARQDLELCFIQEMVSAWQTTFEKRMGLFAWETEVTHPL 1260

Qy     1261  AAYEGCKLVSKPILIAPHLIWLQLLSEMVDTAKYCNRDKVEMFCLLLHRCLPVLKSSKQN 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261  AAYEGCKLVSKPILIAPHLIWLQLLSEMVDTAKYCNRDKVEMFCLLLHRCLPVLKSSKQN 1320

Qy     1321  RQVSTVGCRFKLLQCGLSLLQGNTIPKSLSRNILRERIYSNALDYFCGPPTCPNQSREQL 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321  RQVSTVGCRFKLLQCGLSLLQGNTIPKSLSRNILRERIYSNALDYFCGPPTCPNQSREQL 1380

Qy     1381  LEDIMILLKFWQTMRSEKKHLVTSEVGDYDLTNASVSSTQMLAVRNNPETASLISGGGLV 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381  LEDIMILLKFWQTMRSEKKHLVTSEVGDYDLTNASVSSTQMLAVRNNPETASLISGGGLV 1440

Qy     1441  NDYTRSMSASGNAVGMGMGVAGGGSSSGWYNTIPHSTSTLSKRSNRSKRLQYQKDSYDKD 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441  NDYTRSMSASGNAVGMGMGVAGGGSSSGWYNTIPHSTSTLSKRSNRSKRLQYQKDSYDKD 1500

Qy     1501  YMKKRNLILELLAVELEFLITWYNPNSLPDLIVPGEEQITEWRNRPYKSTVWRDYARLAW 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501  YMKKRNLILELLAVELEFLITWYNPNSLPDLIVPGEEQITEWRNRPYKSTVWRDYARLAW 1560
```

```
Qy   1561 CYNPALAVFLPQRIKNAEIIDEEVSRLVCSDPIAVCHIPEALKYLCTTKNLLQESPDLVY 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 CYNPALAVFLPQRIKNAEIIDEEVSRLVCSDPIAVCHIPEALKYLCTTKNLLQESPDLVY 1620

Qy   1621 ILSWSPVTPIQALAYFSRQYPSHPLTAQYAVKTLSSYPAESVLPYIPQLVQALRHDTMGY 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ILSWSPVTPIQALAYFSRQYPSHPLTAQYAVKTLSSYPAESVLPYIPQLVQALRHDTMGY 1680

Qy   1681 VVEFIKNISRRSQIVAHQLIWNMQTNMYMDEDQQHKDPNLYEALDQLSQSIIASFSGAAK 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 VVEFIKNISRRSQIVAHQLIWNMQTNMYMDEDQQHKDPNLYEALDQLSQSIIASFSGAAK 1740

Qy   1741 RFYEREFDFFGKITAVSGEIRSFAKGIERKNACLAALSRIKVQGGCYLPSNPEAMVLDID 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 RFYEREFDFFGKITAVSGEIRSFAKGIERKNACLAALSRIKVQGGCYLPSNPEAMVLDID 1800

Qy   1801 YSSGTPMQSAAKAPYLARFRVYRCGITELETRAMEVSNNPNSQEDAKMTLGVESWQAAIF 1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 YSSGTPMQSAAKAPYLARFRVYRCGITELETRAMEVSNNPNSQEDAKMTLGVESWQAAIF 1860

Qy   1861 KVGDDVRQDMLALQVITIFKNIFQQVGLDLFLFPYRVVATAPGCGVIECVPNAKSRDQLG 1920
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 KVGDDVRQDMLALQVITIFKNIFQQVGLDLFLFPYRVVATAPGCGVIECVPNAKSRDQLG 1920

Qy   1921 RQTDSGLSEYFQHQYGDESSKEFQAARANFVKSMAAYSLIGYLLQIKDRHNGNIMIDKDG 1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 RQTDSGLSEYFQHQYGDESSKEFQAARANFVKSMAAYSLIGYLLQIKDRHNGNIMIDKDG 1980

Qy   1981 HIIHIDFGFMFESSPGGNIGFEPDMKLTDEMVMIMGGKMDSPAFKWFCELCVQAFLAVRP 2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 HIIHIDFGFMFESSPGGNIGFEPDMKLTDEMVMIMGGKMDSPAFKWFCELCVQAFLAVRP 2040

Qy   2041 YQDAIVSLVSLMLDTGLPCFRGQTINLLKQRFVATKNNKEAAAHMLAVIRNSYQNFRTRT 2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 YQDAIVSLVSLMLDTGLPCFRGQTINLLKQRFVATKNNKEAAAHMLAVIRNSYQNFRTRT 2100

Qy   2101 YDMIQYYQNQIPY 2113
          |||||||||||||
Db   2101 YDMIQYYQNQIPY 2113
```